

RESEARCH

Open Access



Test-retest reliability of physical activity questionnaires in Parkinson's disease

Samuel Ånfors^{1,2*}, Ann-Sofi Kammerlind³ and Maria H. Nilsson^{2,4,5}

Abstract

Background: People with Parkinson's disease are less physically active than controls. It is important to promote physical activity, which can be assessed using different methods. Subjective measures include physical activity questionnaires, which are easy and cheap to administer in clinical practice. Knowledge of the psychometric properties of physical activity questionnaires for people with Parkinson's disease is limited. The aim of this study was to evaluate the test-retest reliability of physical activity questionnaires in individuals with Parkinson's disease without cognitive impairment.

Methods: Forty-nine individuals with Parkinson's disease without cognitive impairment participated in a test-retest reliability study. At two outpatient visits 8 days apart, the participants completed comprehensive questionnaires and single-item questions: International Physical Activity Questionnaire-Short Form (IPAQ-SF), Physical Activity Scale for the Elderly (PASE), Saltin-Grimby Physical Activity Level Scale (SGPALS) and Health on Equal Terms (HOET). Test-retest reliability was evaluated using the intraclass correlation coefficient (ICC), standard error of measurement (SEM), limits of agreement, weighted kappa or the Svensson method.

Results: Several of the physical activity questionnaires had relatively low test-retest reliability, including the comprehensive questionnaires (IPAQ-SF and PASE). Total physical activity according to IPAQ-SF had an ICC value of 0.46 (95% confidence interval [CI], 0.21–0.66) and SEM was 2891 MET-min/week. The PASE total score had an ICC value of 0.66 (95% CI, 0.46–0.79), whereas the SEM was 30 points. The single-item scales of SGPALS-past six months (SGPALS-6 m) and HOET question 1 (HOET-q1) with longer time frames (6 or 12 months, respectively) showed better results. Weighted kappa values were 0.64 (95% CI, 0.45–0.83) for SGPALS-6 m and 0.60 (95% CI, 0.39–0.80) for HOET-q1, whereas the single-item questions with a shorter recall period had kappa values < 0.40.

Conclusions: Single-item questions with a longer time frame (6 or 12 months) for physical activity were shown to be more reliable than multi-item questionnaires such as the IPAQ-SF and PASE in individuals with Parkinson's disease without cognitive impairments. There is a need to develop a core outcome set to measure physical activity in people with Parkinson's disease, and there might be a need to develop new physical activity questionnaires.

Keywords: Parkinson disease, Exercise, Reproducibility of results, Surveys and questionnaires

Background

People with Parkinson's disease (PD) are less physically active and spend more time in sedentary activities than controls [1, 2], even at the early stages of the disease [2]. In comparison with controls, fewer people with PD [2, 3] reach the World Health Organization's global health recommendations of 150 min of moderate- to vigorous-intensity physical activity (PA) per week [4]. Disease

*Correspondence: samuel.anfors@rjl.se

¹ Department of Rehabilitation Medicine, Jönköping, Region Jönköping County, and Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden
Full list of author information is available at the end of the article



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

severity, walking ability, and disability in daily living explain most of the decreased PA in PD, but falls, fear of falling, comorbidity, and depression are also associated with less PA [1]. Low outcome expectation, lack of time, fear of falling, non-motor symptoms, and lack of support are examples of barriers to exercise [5]. According to the European guidelines for physiotherapists who treat people with PD, a key goal is to prevent inactivity and promote PA [6].

PA is defined as “any bodily movement produced by skeletal muscles that results in energy expenditure” [7]. Exercise is PA that is planned, structured, and aims to improve or maintain physical fitness [7]. There is a dose-response relationship between the level of PA and the risk for developing PD [8]. Whether PA alters the disease prognosis remains to be shown, but the health benefits are undoubted [9]. For example, exercise improves muscle strength, balance, and motor symptoms in people with PD [10, 11]. It is important to investigate the level of PA in order to prevent physical inactivity and evaluate the effect of PA interventions.

PA can be assessed using criteria, objective and subjective measures. Criteria measures include measuring heat production, oxygen consumption, or production of carbon dioxide. Such measures are not feasible in clinical practice. Examples of objective measures include accelerometers, pedometers, and heart frequency measurements. Unlike pedometers, accelerometers can measure the intensity of PA. Accelerometers are commonly used in research. Nevertheless, they have limitations in measuring PA without acceleration (e.g., bicycling), activities with higher energy consumption (e.g., climbing stairs), and when PA is performed in water such as swimming [12]. Examples of subjective measures are PA questionnaires (PAQs) and PA diaries. PAQs are suitable for use in larger epidemiologic studies and are commonly used in clinical practice because they are easier and cheaper to administer than objective measures [12, 13]. For a comprehensive assessment of total PA, all domains of PA should be included in the questionnaire such as sports and household activities [14]. Knowledge of the psychometric properties of PAQs for people with PD is limited. For example, no specific PAQ is recommended in the European physiotherapy guidelines for PD [6]. Reliable and valid measures are required for use in clinical practice and research. Moreover, the psychometric properties of an instrument are sample dependent [15].

Methods

The aim of this study was to evaluate the test-retest reliability of PAQs in people with PD without cognitive impairment. The evaluation included both

comprehensive questionnaires and single-item questions: International Physical Activity Questionnaire-Short Form (IPAQ-SF), Physical Activity Scale for the Elderly (PASE), Saltin-Grimby Physical Activity Level Scale (SGPALS), and Health on Equal Terms (HOET).

Participants

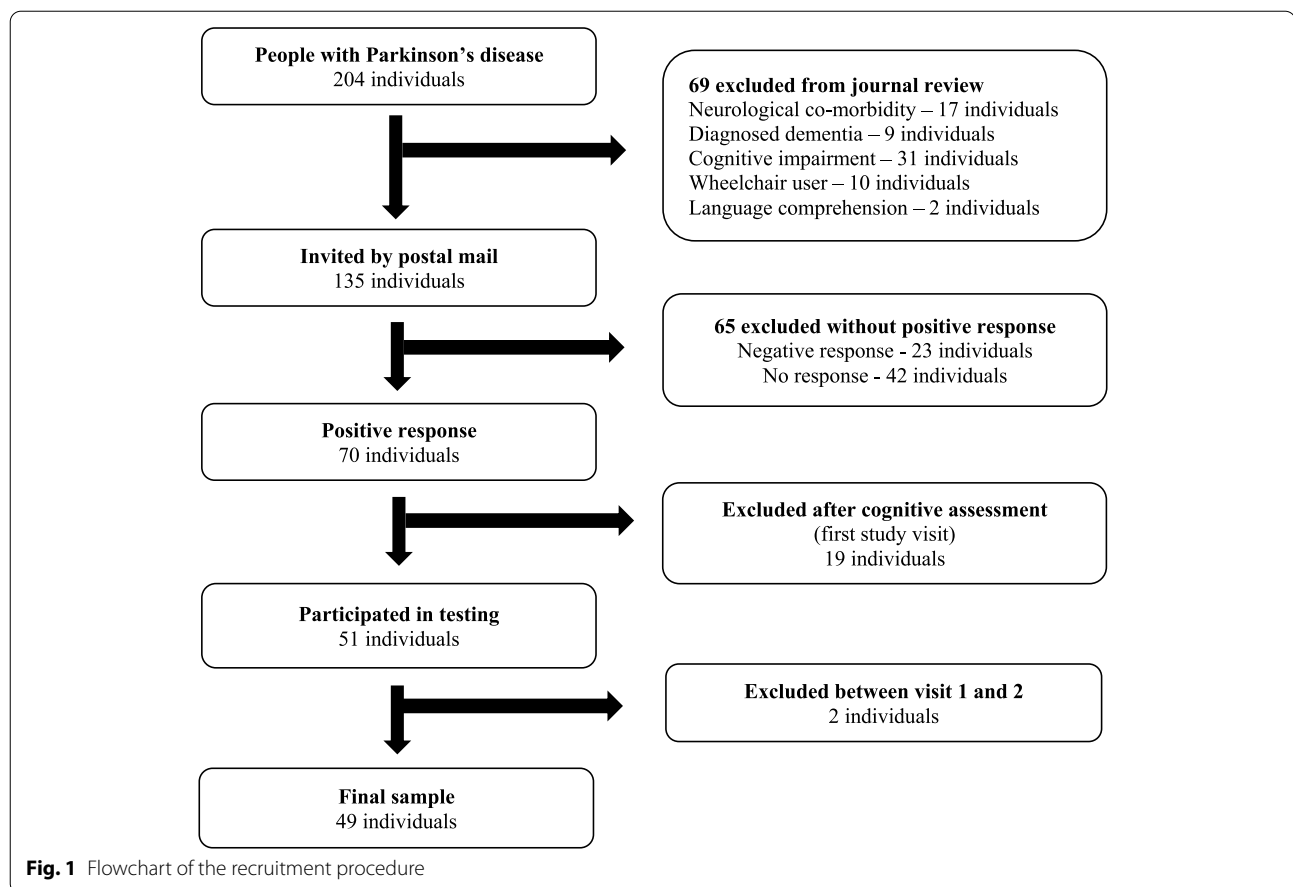
All patients diagnosed with idiopathic PD living in Region Jönköping County, Sweden, who received care at the Internal Medicine or Geriatric departments at County Hospital Ryhov, Jönköping, Sweden, were considered eligible for inclusion. Exclusion criteria were other neurologic disorders, dementia diagnosis, documented cognitive impairments (e.g. mild cognitive disorder or subject to cognitive investigation), using a wheelchair indoors, or insufficient understanding of the Swedish language. The sole use of patient-reported outcome measures among cognitively impaired respondents may introduce additional challenges [16], therefore those who scored <26 points on the Montreal Cognitive Assessment (MoCA) [17] at the first study visit were also excluded. MoCA is a valid instrument for detecting mild cognitive impairment and dementia in individuals with PD [18]. Participants with other co-morbidities were not excluded.

Recruitment procedure

Potential participants were identified using an existing database for patients diagnosed with PD in Region Jönköping County and by screening medical records. The recruitment procedure is presented in Fig. 1. In total, 204 individuals diagnosed with PD were identified, and 69 (mean age, 71 years; standard deviation [SD], 10.1 years; 67% men) were excluded according to the study criteria (Fig. 1). The remaining 135 people with PD received information about the study, a letter of inquiry, and a pre-stamped return envelope by post.

Ninety-three people responded (response rate, 69%). The mean age of those who did not respond ($n=42$) or declined to participate ($n=23$) was 68 years (SD, 12.3 years) and a median duration of PD of 4.4 years (q1–q3, 1.4–8.0 years); 51% were men. The mean age of the 70 participants (63% men) who accepted the invitation was 67 years (SD, 7.7 years) and the median duration of PD was 4.3 years (q1–q3, 1.7–8.8 years). There were no significant differences ($P \geq 0.15$) in relation to sex, age, or duration of PD between those who did not respond or declined versus those who accepted the invitation.

Those who agreed to participate were contacted by phone; they received further information about the study and were invited to an outpatient visit that included cognitive screening. An additional 19 people (78% men)



were excluded because their MoCA score was <26 ; their median age was 71 years (q1–q3, 65–77 years) and their median duration of PD was 3.3 years (q1–q3, 1.4–9.6 years). After the initial visit, one participant withdrew from the study and another sustained a fracture and could not participate. This yielded a final sample of 49 participants: 27 men (55%) and 22 women (45%). The mean age was 65 years (SD, 6.9 years) and a median duration of PD 4.3 years (q1–q3, 1.9–7.1 years). The participants' characteristics are presented in Table 1.

Data collection procedure

The participants were booked for two outpatient visits 8 days apart (median, 8.0 days; min–max, 8–10 days). All participants were examined by the same assessor (SÅ). At both visits, the following PAQs were self-administered in the following order: two versions of SGPALS [19, 20]; two questions from HOET [21, 22]; IPAQ-SF [23], and PASE [24]. SGPALS was administered in two versions using two different time frames for retrospective recall: SGPALS 1 week (SGPALS-w) and SGPALS 6 months (SGPALS-6 m).

At the first visit, the participants were given a booklet of self-administered questions for descriptive purposes. In addition, clinical assessments addressed global cognitive function, motor symptoms, disease severity, and physical capacity. For descriptive purposes and between the two visits, the participants wore an accelerometer: ActiGraph GT1M (ActiGraph LLC, Pensacola, FL, USA) [25], which provides uni-axial measurements of daily PA. The participants wore the accelerometer (placed on the right hip) during waking hours for 7 days, except for activities in water. The accelerometer data are presented in Table 1.

At the second visit, the participants responded to all four PAQs. In order to investigate whether the participants were stable at retest [26], they were given questions on whether they had undergone any changes in medication or deep brain stimulation, answered with yes/no. They also answered questions about changes since their first visit in relation to their walking ability, physical capacity, and/or PA. These questions had five response options: “much better/higher”, “better/higher”, “unchanged”, “worse/lower”, or “much worse/lower”.

Table 1 Characteristics and descriptive information for the sample (N = 49)

Sex; men/women, n (%)	27 (55)/22 (45)
Age (years), mean (SD)	65 (7)
PD duration (years), median (q1–q3)	4.3 (1.9–7.1)
PD severity (H&Y), median (q1–q3)	2 (2–2)
Motor symptoms (UPDRS III), median (q1–q3) ^a	14 (7–19)
Cognitive impairments (MoCA), median (q1–q3)	28 (27–29)
Body mass index (kg/m ²), mean (SD)	26.8 (4.6)
Physical capacity (6-min walk test; meters), mean (SD)	451 (109)
Comorbidity, self-reported (yes), n (%)	22 (45) ^b
Heart and vascular disease, n (%)	10 (20)
Goiter/hypothyroidism, n (%)	3 (6)
Osteoarthritis, n (%)	4 (8)
Pulmonary disease, n (%)	3 (6)
Disease in gastrointestinal or bladder function, n (%)	3 (6)
Diabetes, n (%)	1 (2)
Inguinal hernia, n (%)	1 (2)
Using walking aid indoors, n (%)	5 (10)
Using walking aid outdoors, n (%)	12 (24)
Freezing of gait (scoring ≥ 1 on item 3 of the FOGQsa), n (%)	17 (35)
Falls in the past 6 months, n (%)	10 (20)
Physical activity, accelerometry data ^c	
Sedentary time (min/day), median (q1–q3) ^d	598 (534–670)
Light PA (min/day), median (q1–q3) ^e	166 (112–214)
Lifestyle PA (min/day), median (q1–q3) ^f	31 (21–51)
Moderate PA (min/day), median (q1–q3) ^g	10 (2–20)
Vigorous PA (min/day), median (q1–q3) ^h	0 (0–0)
Very vigorous PA (min/day), median (q1–q3) ⁱ	0 (0–0)

Reported % is valid %

SD standard deviation, PD Parkinson's disease, q1–q3 first and third quartiles, H&Y Hoehn and Yahr staging scale (score range, 1–5; higher = worse), UPDRS III Unified PD Rating Scale, part III – motor examination (score range, 0–108; higher = worse), MoCA the Montreal Cognitive Assessment (score range, 0–30; higher = better), FOGQsa self-administered version of the Freezing of Gait questionnaire, PA physical activity

^a 4 missing values

^b 3 of the 22 participants had > 1 comorbidity

^c n = 46 with valid data

^d 0–99 cpm

^e 100–759 cpm

^f 760–1951 cpm

^g 1952–5724 cpm

^h 5725–9498 cpm

ⁱ > 9498 cpm

Physical activity questionnaires

IPAQ-SF

IPAQ-SF is a self-administered questionnaire, which was created to enable international comparisons of PA [23]. It includes 7 items that cover daily PAs with regard to transportation, work, and household and leisure time. The

questions refer to the past 7 days. The following data were registered in relation to vigorous and moderate activities as well as for walking and sedentary time: the number of days that these activities took place and the mean time per day. A sum of metabolic equivalent task (MET) minutes/week was calculated and the respondents were categorized into groups of low, moderate, or high PA [27]. MET reflects energy expenditure in different PAs [28], and one MET equals the energy consumption at rest. Activities with higher intensity are given higher MET values and when multiplied by frequency and duration, they are presented as MET-minutes per day or per week [29].

IPAQ-SF is one of the most widely used questionnaires on PA, and several studies have evaluated its psychometric properties although not in people with PD [27]. Test-retest reliability of the total PA time (IPAQ-SF) was reported in a study using a pooled Spearman coefficient (0.76). The sample consisted of 1974 generally middle-aged people from 12 countries [23]. Another study of 49 residents in Hong Kong, aged 15–55 years, reported an intraclass correlation coefficient (ICC) of 0.79 [30]. The items on PA level showed inconsistent test-retest reliability results with the lowest ICC value of 0.30 for moderate PA in a group of 108 Norwegian men aged 20–39 years [31]. ICC values for vigorous PA have ranged from 0.61 [31] to 0.75 [30]. ICC values have ranged from 0.80 [31] to 0.97 [30] for sedentary time.

PASE

PASE includes 12 items (ordinal) that cover PA during the past 7 days in three different domains: (1) leisure, (2) household, and (3) work-related activities. It was developed for an elderly population (≥ 65 years) and includes questions on PA with lower energy consumption (e.g., gardening and walking). The different activities are weighted for estimated energy expenditure, with higher weights for more vigorous activities. It yields subscores for the three domains, which are summarized to a total PASE score (range, 0–360 or above). Although PASE includes sedentary activities, these are not included in the total score. Test-retest reliability studies of PASE total score among older people without cognitive impairments reported ICC values ranging from 0.65 to 0.81 [32–34]. ICC values for its subscales have also been reported for older people, although cognitive status was not mentioned: leisure time PA (0.56), household PA (0.94), and work-related PA (0.91) [35]. So far, the psychometric properties of PASE have not been evaluated in people with PD.

SGPALS

SGPALS is a single-item questionnaire; it aims to identify individuals with a sedentary lifestyle and higher risk

profile [36]. It originally covered four levels of PA in relation to sport and leisure [20], but has since been modified to a six-grade scale that also includes domestic activities [19]. The six response categories range from “hardly any physical activity” to “hard or very hard exercise regularly and several times a week, where the physical exertion is great, such as jogging or skiing” [37]. The modified scale was used in this study. Two versions were administered in order to cover different time frames of retrospective recall. The initial question was then revised regarding the time frame: “How much did you move and exert yourself physically during the past week (SGPALS-w)/six months (SGPALS-6 m)?” Test-retest reliability of the six-grade version of SGPALS has been studied in a Finnish project among elderly people, with Pearson correlation coefficients between 0.62 and 0.66 [38, 39].

HOET

Since 2004, the Public Health Agency in Sweden has conducted annual surveys called “Health on Equal Terms” [21]. Two questions concern PA. The first question (HOET-q1) is phrased as follows: “How many times have you exercised and exerted yourself physically in your free time during the past 12 months?” It addresses the leisure time domain and has four response categories (i.e., ordinal scale): “sedentary free time”, “moderate exercise in free time”, “moderate regular exercise in free time”, and “regular exercise and training”. The second question (HOET-q2) is phrased “How many hours in a normal week do you do moderately strenuous activities that make you warm?” This is one of the questions of IPAQ-SF, which has been changed to an ordinal scale with five response alternatives and it concerns “a normal week” instead of the past 7 days. The response alternatives are in reverse order compared with HOET-q1: “5 hours a week or more”, “more than 3 hours but less than 5 hours a week”, “between 1 and 3 hours a week”, “at most 1 hour a week”, and “not at all”. HOET-q2 aims to categorize sufficiently active persons (> 30 min or > 60 min of daily PA on a moderate level) from persons with a sedentary lifestyle [22]. Neither HOET-q1 nor HOET-q2 seems to have been evaluated in relation to test-retest reliability (i.e., regardless of the sample).

Additional descriptive data

Descriptive data included age, sex, and duration of PD (years). Self-administered questions concerned years of education, comorbidity, use of walking aids (indoors and outdoors), a history of falls during the past 6 months and freezing of gait (FOG). FOG was assessed according to item 3 (score 0–4, higher = worse) of the self-administered version [40] of the FOG questionnaire [41] (i.e.

FOGQsa). Those scoring ≥ 1 were categorized as “freezers” [42]. The participant’s weight was measured with a weighing scale (Philips HF 351/00) [43], whereas height was self-reported.

Several clinical assessments were included. Motor symptoms were assessed according to part III of the Unified Parkinson’s Disease Rating Scale (UPDRS) [44]; the total score ranges from 0 to 108 points (higher scores = worse). The severity of PD was assessed according to the Hoehn and Yahr staging scale, which ranges from I to V (higher = worse) [45, 46]. Global cognitive functioning was assessed according to MoCA; the maximal score is 30 points and results ≥ 26 points are considered normal [17]. Physical capacity was tested with the Six-Minute Walk Test [47]; the participants walked for 6 min (fast speed) and the total distance (meters) was measured.

Analysis

Normally distributed data are presented as means and standard deviation, whereas non-normally distributed data are presented as medians and first and third quartiles.

Regarding accelerometry, at least 10 h of wear time a day and 4 valid days were required to get valid data for analysis. This follows recommendations for use in research settings [48, 49]. ActiLife software (ActiGraph LLC, Pensacola, FL, USA) was used to process the data [50]. Data were collected in epochs of 10 s and summarized to 60 s for analysis of the data. There is no consensus on which algorithms to use when analyzing. In accordance with Choi et al. [51], 90 min of no activity was considered nonwear time, allowing a spike tolerance of 2 min with non-zero counts.

Test-retest reliability analyses were done using different methods, depending on the nature of the data and to gain a comprehensive analysis. The ICC, two-way mixed, absolute agreement, [52] was used in relation to the continuous scales of IPAQ-SF and PASE. ICC values were calculated for their respective subscales as well as for the total score. For group-level comparisons, acceptable ICC values should exceed 0.70 [16, 53]. If used for decisions on an individual level, it has been suggested that the ICC value should be at least 0.90 [16]. In addition, to obtain an absolute value of the measurement error, the standard error of measurement (SEM) was calculated using the formula: $SEM = SD_{test} \times \sqrt{1 - ICC}$.

Moreover, limits of agreement (LoA) were visualized with Bland-Altman plots [54] for the differences and means of the paired data in PASE and IPAQ-SF, a method recommended when evaluating questionnaires [55]. In this study, the purpose was to mark the width of the measurement error and detect proportional bias.

Quadratic weighted kappa was used for the ordinal data of SGPALS and HOET. Landis and Koch [56] have proposed the following as standards for strength of agreement for the kappa coefficient: 0.01 = poor, 0.01–0.20 = slight, 0.21–0.40 = fair, 0.41–0.60 = moderate, 0.61–0.80 = substantial, and 0.81–1.00 = almost perfect.

The statistical methods of ICC and kappa are recommended in reliability studies of patient-reported outcome measures [53]. They were complemented by a method described by Svensson [57] (Örebro University, Sweden) for testing the stability of paired ordinal data, which provides a more comprehensive analysis. According to the Svensson method, systematic disagreement and additional individual variations are measured and expressed in terms of percentage agreement (PAgr), relative position (RP), relative concentration (RC), and relative rank variance (RV) [57]. PAgr explains the proportion of accurately matched paired answers with no change over time, values ranging from 0 to 100; a value of 100 indicates full agreement and high values are preferable. RP reflects whether a systematic disagreement between test and retest is present on a group level or not. RC measures any shift in concentration from test to retest, that is, if the paired answers are clustered differently. For RP and RC, the results range from -1 to 1; the desirable value is close to 0. The 95% confidence interval (CI) should include 0 to be considered a stable value for the paired data. For RV, the results can range between 0 and 1, with preferable values close to 0 indicating a homogeneous individual change. High RV values indicate greater individual variance, independent of the systematic disagreement on a group level [57]. The Svensson method was

used for the ordinal data of SGPALS and HOET. $P < 0.05$ was considered a significant difference.

SPSS software, version 20 (SPSS Inc., Chicago, IL, USA.) [58] was used for the statistical analysis. Weighted kappa was analyzed with SAS (SAS Institute Inc., Cary, NC, USA) [59] and Svensson's method with a program provided on an internet web page [60].

Results

No statistically significant differences in PA scores were found between test and retest in any of the PAQs. Six participants (12%) had missing data on SGPALS-6 m; there were no missing data for the remaining PAQs. Moreover, 45 of 49 (92%) participants reported that their walking ability as well as physical capacity was unchanged between test and retest. PA was reported as changed by 11 participants (22%); 2 much higher, 5 higher, and 4 lower. They had no significant differences between test and retest in any of the PAQs examined. Sensitivity analysis was done with these 11 participants excluded, and the results were comparable with no decisive differences in the reliability analysis. Therefore, all 49 participants were included in the study.

Physical activity level

The participants were physically active mostly at a moderate level, according to the assessment with IPAQ-SF and PASE (Table 2). The IPAQ-SF results (test and retest) resulted in categorization of the respondents into low (31 and 24%), moderate (43 and 47%), and high (27 and 29%) PA. Most (63 and 69%) had no vigorous PA at all. According to PASE, they got most of their PA from household-related activities: 68 and 73% in the test and retest,

Table 2 Test-retest reliability of IPAQ-SF and PASE in people with Parkinson's disease (N = 49)

	Test 1		Test 2		ICC (95% CI)	SEM ^a
	Median (q1–q3)	Mean (SD)	Median (q1–q3)	Mean (SD)		
IPAQ-SF						
Vigorous, MET-min/week	0 (0–480)	559 (1268)	0 (0–600)	642 (1309)	0.37 (0.09–0.59)	1010
Moderate, MET-min/week	480 (0–1440)	1593 (2958)	480 (0–1260)	1132 (1663)	0.36 (0.10–0.58)	2360
Walking, MET-min/week	594 (264–1040)	777 (905)	594 (297–1040)	928 (1087)	0.51 (0.27–0.69)	634
Sedentary time, min/week	1890 (1260–2100)	1933 (960)	1680 (1260–2100)	1852 (925)	0.60 (0.38–0.75)	610
Total, MET-min/week	1173 (744–3592)	2930 (3950)	1674 (804–3822)	2702 (2726)	0.46 (0.21–0.66)	2891
PASE						
Leisure time subscore	19 (7–33)	23 (19)	15 (8–28)	24 (27)	0.49 (0.25–0.68)	14
Household subscore	85 (50–106)	78 (40)	85 (50–91)	79 (34)	0.69 (0.51–0.81)	22
Work subscore	0 (0–0)	16 (40)	0 (0–0)	10 (25)	0.30 (0.02–0.53)	33
Total score	118 (81–153)	117 (51)	104 (75–142)	113 (53)	0.66 (0.46–0.79)	30

IPAQ-SF, International Physical Activity Questionnaire, short form; PASE, Physical Activity Scale for the Elderly; q1–q3, first and third quartiles; SD, standard deviation; ICC, two-way mixed intraclass correlation coefficient; CI, confidence interval; SEM, standard error of measurement; MET, metabolic equivalent task

^a SEM defined as $SEM = SD_{test} \times \sqrt{1 - ICC}$

respectively. The ordinal scales of SGPALS and HOET resulted in a concentration of participants categorized in a moderate level of PA (Fig. 2). SGPALS categorized none at level 1 (“hardly any PA”) and very few at level 5–6 (vigorous PA). The results of HOET were better spread but still with nearly half of the participants at a moderate level of PA. The accelerometer data showed even less PA with a dominance of sedentary time, indicating a discrepancy between the PAQs and the accelerometers (Table 1). Valid accelerometer data were obtained from 46 participants (94%).

Test-retest reliability of IPAQ-SF and PASE

The ICC values for the subscales of IPAQ-SF ranged from 0.36 for moderate PA to 0.60 for sedentary time. Total PA had an ICC value of 0.46 and SEM was 2891 MET-min/week (Table 2). Thus, in nearly all aspects, IPAQ-SF presented low test-retest reliability with high measurement error.

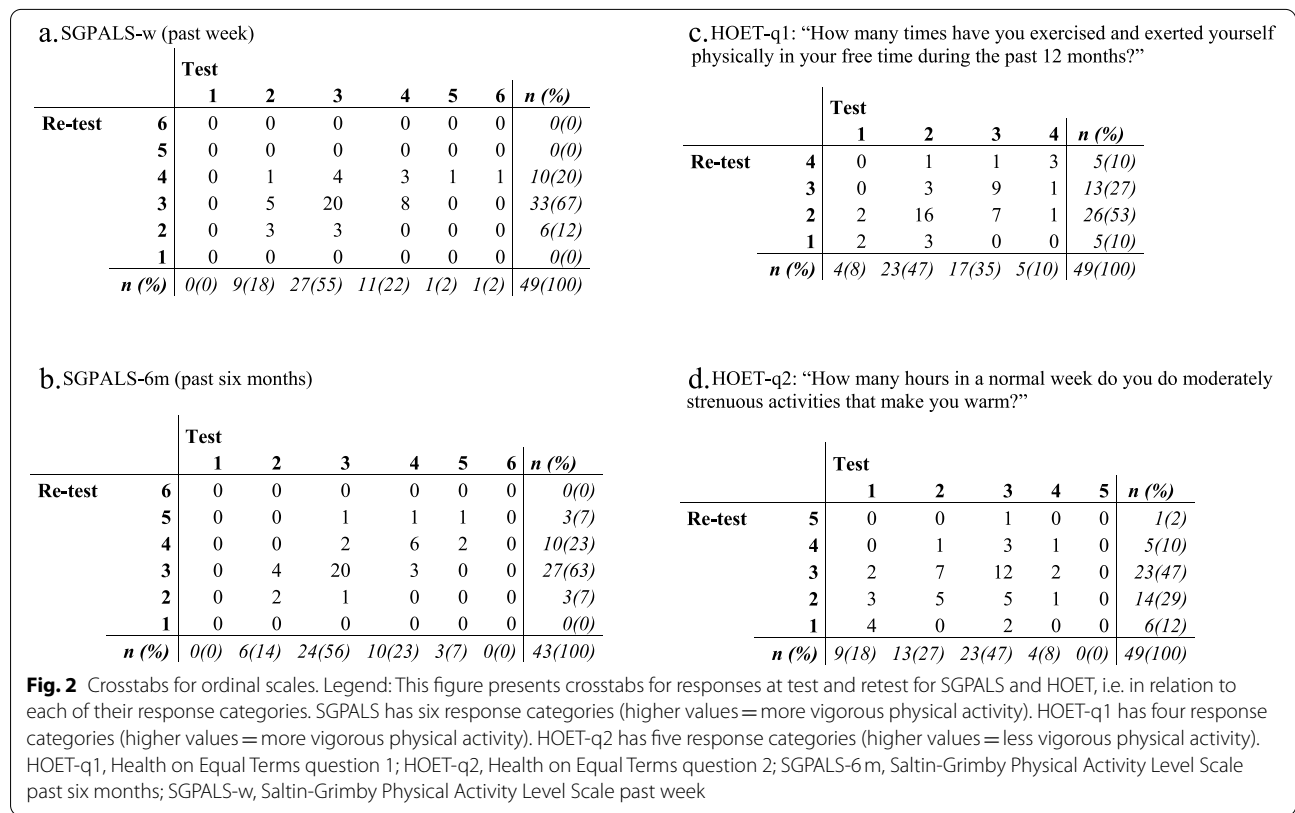
ICC values for the subscales of PASE ranged from 0.30 for work-related PA to 0.69 for household-related PA. The PASE total score had an ICC value of 0.66, whereas the SEM score was 30 points (Table 2).

LoAs are presented with Bland-Altman plots for the differences and means of the paired data in the total scores of IPAQ-SF and PASE (Fig. 3). Both total scales

and their subscales show heteroscedastic values (i.e. the magnitude of differences increases proportionally to the size of the measurement), except for household PA in PASE. For the other comparisons, the difference between test and retest values increased with higher mean values. The mean difference (SD) for total PA in IPAQ-SF was 228 (3526) and LoA – 6684 and 7140 MET-min/week. PASE total score had a mean difference (SD) score of 3 (43) and LoA – 81 and 88. For total PA in IPAQ-SF, 5 outliers (10%) exceeded the LoA, whereas PASE total score had 3 outliers (6%) (Fig. 3).

Test-retest reliability of SGPALS and HOET

The test-retest results of SGPALS and HOET are presented in Table 3. According to weighted kappa values, SGPALS-6m (0.64) and HOET-q1 (0.60) had substantial and moderate agreement, respectively. SGPALS-w and HOET-q2 had weighted kappa values <0.40. The Svensson method showed best PAg_r for SGPALS-6m (67%) followed by HOET-q1 (61%) (Table 3). The RP and RC were close to zero for both versions of SGPALS as well as for both HOET questions, indicating no systematic disagreement on a group level. Moreover, RV results had a CI that included zero except for HOET-q2 (0.16; 95% CI, 0.04–0.29); this is a result of individual variability. The



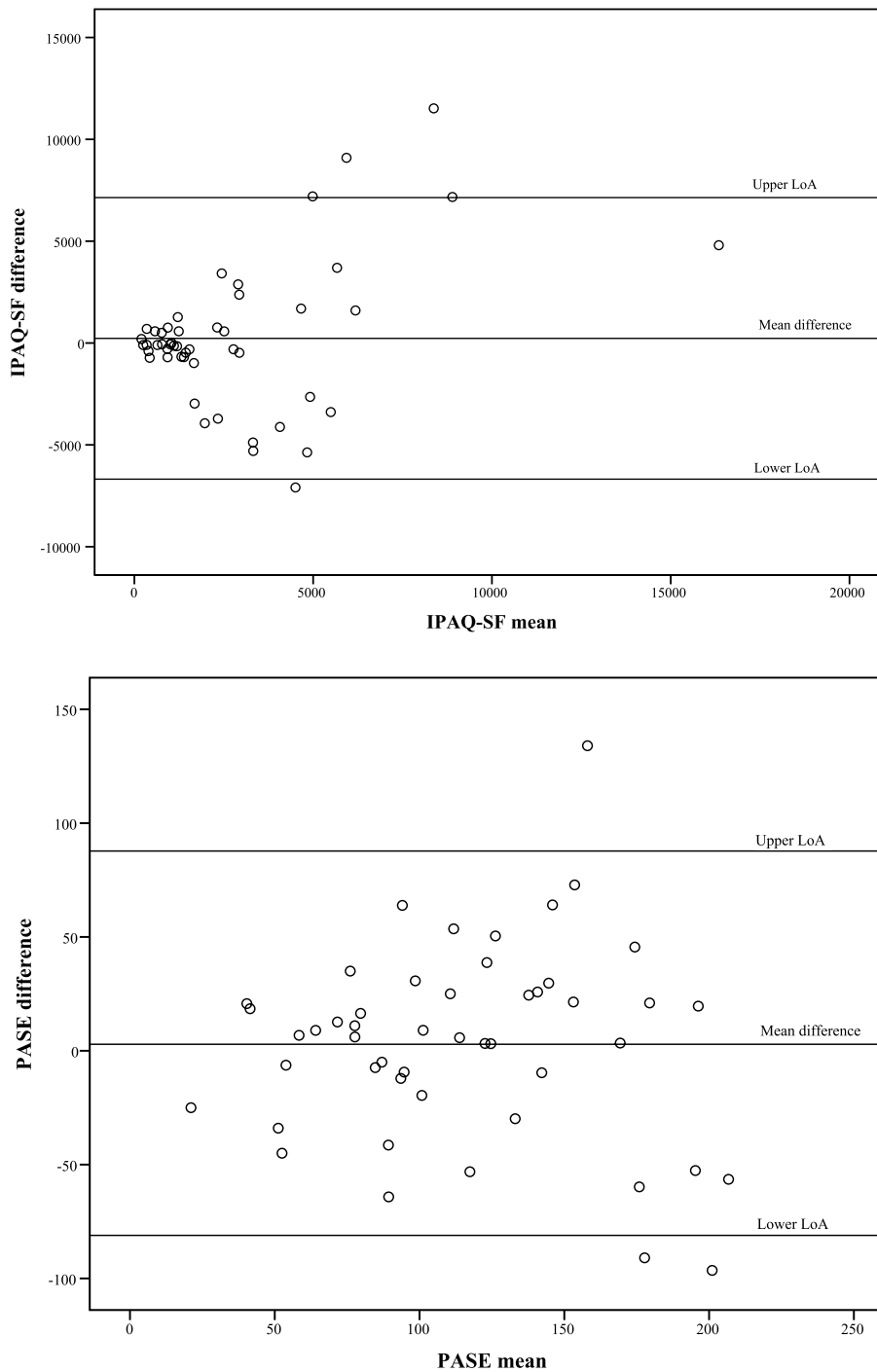


Fig. 3 Limits of agreement with Bland-Altman plots for the total scores for IPAQ-SF and PASE. Legend: Each dot visualizes the difference in two measurements (test and retest; y-axis) from the same individual, and in relation to the average value of the two measurements (x-axis) for this individual. The mid horizontal line is drawn at the mean difference, which is the estimated bias. The lines above and below represent the upper and lower limits of agreement (defined as the mean difference $\pm 1.96 \times$ SD of the differences). The results for IPAQ-SF are presented as the sum of MET-minutes/week (higher values = more vigorous physical activity). The total PASE score can range from 0 to 360 (or above, higher values = more vigorous physical activity). IPAQ-SF, International Physical Activity Questionnaire, short form; LoA, limit of agreement; PASE, Physical Activity Scale for the Elderly; SD, standard deviation

Table 3 Test-retest reliability of SGPALS and HOET (N = 49)

		Test 1: median (q1–q3)	Test 2: median (q1–q3)	wKappa (95% CI)
SGPALS-w ^a		3 (3–4)	3 (3–3)	0.39 (0.18–0.61)
SGPALS-6m ^{a,b}		3 (3–4)	3 (3–4)	0.64 (0.45–0.83)
HOET-q1 ^c		2 (2–3)	2 (2–3)	0.60 (0.39–0.80)
HOET-q2 ^d		3 (2–3)	3 (2–3)	0.39 (0.16–0.62)
	PAgr (%)	RP (95% CI)	RC (95% CI)	RV (95% CI)
SGPALS-w	53	−0.01 (−0.17 to 0.14)	0.13 (−0.01 to 0.26)	0.08 (0.00–0.19)
SGPALS-6 m	67	0.05 (−0.07 to 0.17)	0.08 (−0.07 to 0.22)	0.02 (0.00–0.04)
HOET-q1	61	−0.08 (−0.21 to 0.05)	−0.03 (−0.17 to 0.12)	0.04 (0.00–0.10)
HOET-q2	45	0.08 (−0.08 to 0.25)	0.01 (−0.17 to 0.19)	0.16 (0.04–0.29)

SGPALS, Saltin-Grimby Physical Activity Level Scale; HOET, Health on Equal Terms; wKappa, kappa with quadratic weighting; q1–q3, first and third quartiles; CI, confidence interval; SGPALS-w, SGPALS past week; SGPALS-6 m, SGPALS past six months; HOET-q1, HOET question 1; HOET-q2, HOET question 2; PAgr, percentage agreement; RP, relative position; RC, relative concentration; RV, relative rank variance; PA, physical activity

^a Score 1–6, higher = more PA

^b n = 43 (6 missing)

^c Score 1–4, higher = more PA

^d Score 1–5, higher = less PA

best reliability results were found for SGPALS-6m and HOET-q1.

Discussion

This is the first evaluation of test-retest reliability in people with PD for the PAQs used in this study.

The main results of this study are that several of the PAQs had relatively low test-retest reliability, including the comprehensive questionnaires of IPAQ-SF and PASE. Best results were found for the single-item scales, SGPALS-6m and HOET-q1, which use longer time frames (6 or 12 months).

Test-retest reliability results

IPAQ-SF did not reach acceptable test-retest reliability with lower ICC values than recommended [16, 53], and the SEM value almost exceeded the mean value. Previous studies of a general population aged 15–65 years reported acceptable reliability for total PA (coefficients of 0.76–0.79 [23, 30]), but one of these studies reported Spearman correlations [23]. This discrepancy in findings probably reflects sample differences. The present findings may not be so surprising because IPAQ-SF was developed for younger respondents (18–65 years) [23].

PASE did not reach the limit of acceptable test-retest reliability, although the total PASE score had an ICC value of 0.66, which is close to the recommended cut-off of 0.70 [16, 53]. Because PASE was developed for use with older people ≥65 years of age [24], we had anticipated better results. Previous studies of healthy older persons reported ICC values ranging from 0.79 to 0.99 for PASE total score, and PASE has been recommended for use in groups of older adults [14]. Although test-retest

reliability has been acceptable in different disease-specific samples [61–63], this is the first study to evaluate this in a PD sample. Moreover, the SEM value was 30 points and the LoAs were wide. These findings corroborate previous studies that reported a large measurement error in relation to PASE [34, 61–63]. However, in nearly all aspects, PASE showed better reliability results than IPAQ-SF.

Only a few previous studies of SGPALS and no studies of HOET have evaluated test-retest reliability. Of the PAQs used in this study, only SGPALS-6 m and HOET-q1 had acceptable results for test-retest reliability. Both are single-item questions. This is a surprising finding because scales with multiple items have been suggested to render better reliability results [16]. The present findings are difficult to explain. It might be that single-item questions are easier to comprehend and/or that one should preferably use longer periods for retrospective recall in relation to PA for this population. The latter is further supported by the fact that SGPALS-w and HOET-q2, which both have a 1-week time frame, showed low kappa results.

When choosing PAQs, the purpose must be clear; that is, whether the aim is to categorize individuals according to PA level or whether the aim is to evaluate changes over time or the effect of an intervention. The ordinal scales of SGPALS-6 m and HOET-q1 seem to be suitable for categorizing individuals into groups of different activity levels for research purposes or in clinical settings. Although single item questions are easy to use, it needs to be noted that we lose detailed information as compared to multi-item instruments, and they are less sensitive for detecting changes in types of activity and intensity (e.g. over time or intervention effects) [64]. Our findings suggest that it is preferable to complement PAQs with objective

measures of PA, such as using accelerometers, for example in intervention studies.

Physical activity level and future perspectives

Our data showed low levels of PA in people with PD, consistent with previous studies [1–3]. However, one study showed that PASE scores did not differ significantly from controls when only including participants with an early PD (i.e. mean PD duration was 16.6 months) [65]. This underlines the importance of focusing on maintaining PA after diagnosis.

The participants in our study were relatively young with mild PD without cognitive impairments, and they still had a sedentary lifestyle. This is important knowledge because promoting PA is one of the key components for physical therapists who treat individuals with PD [6]. For assessment of PA in people with PD, it is important to choose measures that include low energy PA, because the participants were physically active mostly at a moderate level. Moreover, they got most of their PA from household-related activities, which underlines the importance of including such activities when evaluating PA. We recommend developing a core outcome set [66] for measuring PA in people with PD, with a consensus on what concepts of PA to assess and how to assess it. There might even be a need to develop new subjective measures.

Methodological considerations: strengths and limitations

This study addresses the paucity of psychometric studies in relation to PAQs within the field of PD, and the knowledge gained is anticipated to be important for both clinicians and researchers. A strength is that test-retest reliability was comprehensively investigated in relation to several PAQs. Moreover, the final sample size of 49 participants is close to the recommendation of 50 participants for reliability studies [67, 68]. However, we do acknowledge that other psychometric properties are also important.

Participants with cognitive impairments were excluded, which affects the external validity of the present findings. That is, our findings are not valid for people with PD with cognitive impairments, and mild cognitive impairment is common in those with PD [69]. The reasoning for excluding people with cognitive impairments was that subjective measures and patient-reported outcomes require good cognitive functioning for reliable results due to retrospective recall [70]. Therefore, as an initial step, we focused on evaluating test-retest reliability of PAQs in individuals with PD without cognitive impairments.

Eight days were used between test and retest, which implies that two different weeks were the objects of the PAQs. This might give biased results because PA varies from day to day, between weeks, and during

seasons of the year [71–73]. However, we detected no statistically significant difference in PA between test and retest. A shorter period between test and retest has been associated with a higher reliability coefficient [74]. If the time span is too long, a greater risk for an actual change occurs. At retest and as recommended [26], specific questions addressed whether the participant had undergone or perceived any changes since the first administration of the PAQs. Eleven participants reported a change in PA between test and retest (no statistical differences in PAQ scores): 7 reported a higher PA level and 4 a lower PA level. The fact that more participants reported a higher PA level at retest might mirror that they had been using accelerometers, possible motivators to increase their PA. This should be kept in mind in future studies.

Participants with PD can have both motor and non-motor fluctuations. Although we did include specific questions that addressed whether the participant had undergone or perceived any changes since the first administration of the PAQs, it would have been of value to include descriptive data of both motor and non-motor symptoms also at retest.

Conclusions

Single-item questions with a longer time frame (6 or 12 months) for PA were shown to be more reliable than multi-item questionnaires such as the IPAQ-SF and PASE in people with PD without cognitive impairments. There is a need to develop a core outcome set for measuring PA in people with PD, and there might be a need to develop new PAQs.

Abbreviations

CI: Confidence interval; FOG: Freezing of gait; FOGQsa: the self-administered version of the FOG questionnaire; HOET: Health on Equal Terms; HOET-q1: HOET question 1; HOET-q2: HOET question 2; ICC: Intraclass correlation coefficient; IPAQ-SF: International Physical Activity Questionnaire-Short Form; LoA: Limit of agreement; MET: Metabolic equivalent task; MoCa: Montreal Cognitive Assessment; PA: Physical activity; PAgr: Percentage agreement; PAQ: PA questionnaire; PASE: Physical Activity Scale for the Elderly; PD: Parkinson's disease; RC: Relative concentration; RP: Relative position; RV: Relative rank variance; SD: Standard deviation; SEM: Standard error of measurement; SGPALS: Saltin-Grimby Physical Activity Level Scale; SGPALS-6m: SGPALS past six months; SGPALS-w: SGPALS past week; UPDRS: Unified Parkinson's Disease Rating Scale.

Acknowledgments

The authors wish to acknowledge Helene Jacobsson, Bo Rolander, and Mats Nilsson for their support on statistical methods, Åsa Tornberg for accelerometer support, and all the people with PD who participated in the study.

Authors' contributions

SÅ and MHN designed the study, and all authors participated in the ethical application. SÅ collected all the data and analyzed the data in collaboration with MHN. SÅ drafted the initial manuscript, which was repeatedly critically revised by MHN. AK provided input on design as well as the manuscript. All authors read and approved the final manuscript.

Funding

MHN was funded by the Strategic Research Area in neuroscience at Lund University, Sweden (MultiPark). SÅ was funded by (1) Futurum, Academy for Health and Care, Region Jönköping County, Sweden, (2) the Swedish Parkinson Academy, Sweden, and (3) the Swedish Association of Physiotherapists, Sweden.

Availability of data and materials

All relevant data are within the manuscript. The dataset generated and analyzed in the current study is not publicly available due to privacy constraints relating to the ethical approval and informed consent signed by the participants. Data sharing was not stated in the informed consent signed by the participants.

Declarations

Ethics approval and consent to participate

The study was approved by the Regional Ethical Review Board in Linköping, Sweden (no. 2011/373–31); all participants provided written informed consent and the study was performed in accordance with the Declaration of Helsinki guidelines.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Rehabilitation Medicine, Jönköping, Region Jönköping County, and Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden. ²Department of Health Sciences, Lund University, Lund, Sweden. ³Futurum, Region Jönköping County, and Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden. ⁴Memory Clinic, Skåne University Hospital, Malmö, Sweden. ⁵Clinical Memory Research Unit, Department of Clinical Sciences Malmö, Lund University, Lund, Sweden.

Received: 17 June 2021 Accepted: 22 September 2021

Published online: 15 October 2021

References

- van Nimwegen M, Speelman AD, Hofman-van Rossum EJ, Overeem S, Deeg DJ, Borm GF, et al. Physical inactivity in Parkinson's disease. *J Neurol*. 2011;258(12):2214–21. <https://doi.org/10.1007/s00415-011-6097-7>.
- Lord S, Godfrey A, Galna B, Mhiripiri D, Burn D, Rochester L. Ambulatory activity in incident Parkinson's: more than meets the eye? *J Neurol*. 2013;260(12):2964–72. <https://doi.org/10.1007/s00415-013-7037-5>.
- Benka Wallen M, Franzen E, Nero H, Hagstromer M. Levels and patterns of physical activity and sedentary behavior in elderly people with mild to moderate Parkinson disease. *Phys Ther*. 2015;95(8):1135–41. <https://doi.org/10.2522/ptj.20140374>.
- World Health Organization. Global recommendations on physical activity for health. Geneva: World Health Organization; 2010.
- Schootemeijer S, van der Kolk NM, Ellis T, Mirelman A, Nieuwboer A, Nieuwhof F, et al. Barriers and motivators to engage in exercise for persons with Parkinson's disease. *J Parkinsons Dis*. 2020;10(4):1293–9. <https://doi.org/10.3233/jpd-202247>.
- Keus SHJ, Munneke M, Graziano M; on behalf of the guideline development group. European physiotherapy guideline for Parkinson's disease 2014. <https://www.parkinsonnet.com/discipline/physiotherapy/>. Accessed 31 Aug 2021.
- Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Rep*. 1985;100(2):126–31.
- Fang X, Han D, Cheng Q, Zhang P, Zhao C, Min J, et al. Association of levels of physical activity with risk of Parkinson disease: a systematic review and meta-analysis. *JAMA Netw Open*. 2018;1(5):e182421. <https://doi.org/10.1001/jamanetworkopen.2018.2421>.
- Marras C, Canning CG, Goldman SM. Environment, lifestyle, and Parkinson's disease: implications for prevention in the next decade. *Mov Disord*. 2019;34(6):801–11. <https://doi.org/10.1002/mds.27720>.
- Li X, He J, Yun J, Qin H. Lower limb resistance training in individuals with Parkinson's disease: an updated systematic review and meta-analysis of randomized controlled trials. *Front Neurol*. 2020;11:591605. <https://doi.org/10.3389/fneur.2020.591605>.
- Radder DLM, Lígia Silva de Lima A, Domingos J, SHJ K, van Nimwegen M, Bloem BR, et al. Physiotherapy in Parkinson's disease: a meta-analysis of present treatment modalities. *Neurorehabil Neural Repair*. 2020;34(10):871–80. <https://doi.org/10.1177/1545968320952799>.
- Vanhees L, Lefevre J, Philippaerts R, Martens M, Huygens W, Troosters T, et al. How to assess physical activity? How to assess physical fitness? *Eur J Cardiovasc Prev Rehabil*. 2005;12(2):102–14. <https://doi.org/10.1097/01.hjr.0000161551.73095.9c>.
- Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1.9 million participants. *Lancet Glob Health*. 2018;6(10):e1077–86. [https://doi.org/10.1016/s2214-109x\(18\)30357-7](https://doi.org/10.1016/s2214-109x(18)30357-7).
- Sattler MC, Jaunig J, Tösch C, Watson ED, Morkink LB, Dietz P, et al. Current evidence of measurement properties of physical activity questionnaires for older adults: an updated systematic review. *Sports Med*. 2020;50(7):1271–315. <https://doi.org/10.1007/s40279-020-01268-x>.
- Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess*. 2009;13(12):iii, ix–x, 1–177. <https://doi.org/10.3310/hta13120>.
- Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value Health*. 2007;10(Suppl 2):S94–105. <https://doi.org/10.1111/j.1524-4733.2007.00272.x>.
- Nasreddine ZS, Phillips NA, Bedirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 2005;53(4):695–9. <https://doi.org/10.1111/j.1532-5415.2005.53221.x>.
- Hoops S, Nazem S, Siderowf AD, Duda JE, Xie SX, Stern MB, et al. Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson disease. *Neurology*. 2009;73(21):1738–45. <https://doi.org/10.1212/WNL.0b013e3181c34b47>.
- Mattiasson-Nilo I, Sonn U, Johannesson K, Gosman-Hedstrom G, Persson GB, Grimby G. Domestic activities and walking in the elderly: evaluation from a 30-hour heart rate recording. *Aging (Milano)*. 1990;2(2):191–8. <https://doi.org/10.1007/BF03323916>.
- Saltin B, Grimby G. Physiological analysis of middle-aged and old former athletes. Comparison with still active athletes of the same ages. *Circulation*. 1968;38(6):1104–15.
- Nationella folkhälsoenkäten - Hälsa på lika villkor. <https://www.folkhalsomyndigheten.se/folkhalsorapportering-statistik/om-vara-datainsamlningar/nationella-folkhalsoenkaten/>. Accessed 31 Aug 2021.
- Boström G. Objective and background of the questions in the national public health survey. Stockholm: Statens folkhälsoinstitut; 2009.
- Craig CL, Marshall AL, Sjostrom M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc*. 2003;35(8):1381–95. <https://doi.org/10.1249/01.MSS.0000078924.61453.FB>.
- Washburn RA, Smith KW, Jette AM, Janney CA. The physical activity scale for the elderly (PASE): development and evaluation. *J Clin Epidemiol*. 1993;46(2):153–62. [https://doi.org/10.1016/0895-4356\(93\)90053-4](https://doi.org/10.1016/0895-4356(93)90053-4).
- ActiGraph. <http://actigraphcorp.com/>. Accessed 31 Aug 2021.
- Terwee CB, Morkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*. 2012;21(4):651–7. <https://doi.org/10.1007/s11136-011-9960-1>.
- International Physical Activity Questionnaire. <http://www.ipaq.ki.se/>. Accessed 31 Aug 2021.
- Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, et al. Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc*. 2000;32(9 Suppl):S498–504. <https://doi.org/10.1097/00005768-200009001-00009>.

29. Hagstromer M, Oja P, Sjostrom M. The international physical activity questionnaire (IPAQ): a study of concurrent and construct validity. *Public Health Nutr.* 2006;9(6):755–62. <https://doi.org/10.1079/PHN2005898>.
30. Macfarlane DJ, Lee CC, Ho EY, Chan KL, Chan DT. Reliability and validity of the Chinese version of IPAQ (short, last 7 days). *J Sci Med Sport.* 2007;10(1):45–51. <https://doi.org/10.1016/j.jsams.2006.05.003>.
31. Kurtze N, Rangul V, Hustvedt BE. Reliability and validity of the international physical activity questionnaire in the Nord-Trøndelag health study (HUNT) population of men. *BMC Med Res Methodol.* 2008;8:63. <https://doi.org/10.1186/1471-2288-8-63>.
32. Hagiwara A, Ito N, Sawai K, Kazuma K. Validity and reliability of the physical activity scale for the elderly (PASE) in Japanese elderly people. *Geriatr Gerontol Int.* 2008;8(3):143–51. <https://doi.org/10.1111/j.1447-0594.2008.00463.x>.
33. Ngai SP, Cheung RT, Lam PL, Chiu JK, Fung EY. Validation and reliability of the physical activity scale for the elderly in Chinese population. *J Rehabil Med.* 2012;44(5):462–5. <https://doi.org/10.2340/16501977-0953>.
34. Vaughan K, Miller WC. Validity and reliability of the Chinese translation of the physical activity scale for the elderly (PASE). *Disabil Rehabil.* 2013;35(3):191–7. <https://doi.org/10.3109/09638288.2012.690498>.
35. Dinger MK, Oman RF, Taylor EL, Vesely SK, Able J. Stability and convergent validity of the physical activity scale for the elderly (PASE). *J Sports Med Phys Fitness.* 2004;44(2):186–92.
36. Rodjer L, Jonsdottir IH, Rosengren A, Bjorck L, Grimby G, Thelle DS, et al. Self-reported leisure time physical activity: a useful assessment tool in everyday health care. *BMC Public Health.* 2012;12:693. <https://doi.org/10.1186/1471-2458-12-693>.
37. Frändin K, Grimby G. Assessment of physical activity, fitness and performance in 76-year-olds. *Scand J Med Sci Sports.* 1994;4(1):41–6. <https://doi.org/10.1111/j.1600-0838.1994.tb00404.x>.
38. Rantanen T, Era P, Heikkinen E. Physical activity and the changes in maximal isometric strength in men and women from the age of 75 to 80 years. *J Am Geriatr Soc.* 1997;45(12):1439–45. <https://doi.org/10.1111/j.1532-5415.1997.tb03193.x>.
39. Sihvonen S, Rantanen T, Heikkinen E. Physical activity and survival in elderly people: a five-year follow-up study. *J Aging Phys Act.* 1998;6(2):133–144. <https://doi.org/10.1123/japa.6.2.133>.
40. Nilsson MH, Hariz GM, Victorin K, Miller M, Forsgren L, Hagell P. Development and testing of a self administered version of the freezing of gait questionnaire. *BMC Neurol.* 2010;10:85. <https://doi.org/10.1186/1471-2377-10-85>.
41. Giladi N, Shabtai H, Simon ES, Biran S, Tal J, Korczyn AD. Construction of freezing of gait questionnaire for patients with parkinsonism. *Parkinsonism Relat Disord.* 2000;6(3):165–70. [https://doi.org/10.1016/s1353-8020\(99\)00062-0](https://doi.org/10.1016/s1353-8020(99)00062-0).
42. Nilsson MH, Hariz GM, Iwarsson S, Hagell P. Walking ability is a major contributor to fear of falling in people with Parkinson's disease: implications for rehabilitation. *Parkinsons Dis.* 2012;2012:713236. <https://doi.org/10.1155/2012/713236>.
43. Weighing scales standard HF351/00. http://www.philips.se/c-p/HF351_00/. Accessed 31 Aug 2021.
44. MC FS, Calne D, Goldstein M. Recent developments in Parkinson's disease. Florham Park, NJ: Macmillan Healthcare Information; 1987.
45. Goetz CG, Poewe W, Rascol O, Sampaio C, Stebbins GT, Counsell C, et al. Movement Disorder Society task force report on the Hoehn and Yahr staging scale: status and recommendations. *Mov Disord.* 2004;19(9):1020–8. <https://doi.org/10.1002/mds.20213>.
46. Hoehn MM, Yahr MD. Parkinsonism: onset, progression and mortality. *Neurology.* 1967;17(5):427–42. <https://doi.org/10.1212/wnl.17.5.427>.
47. Guyatt GH, Sullivan MJ, Thompson PJ, Fallen EL, Pugsley SO, Taylor DW, et al. The 6-minute walk: a new measure of exercise capacity in patients with chronic heart failure. *Can Med Assoc J.* 1985;132(8):919–23.
48. Ward DS, Evenson KR, Vaughn A, Rodgers AB, Troiano RP. Accelerometer use in physical activity: best practices and research recommendations. *Med Sci Sports Exerc.* 2005;37(11 Suppl):S582–8. <https://doi.org/10.1249/01.mss.0000185292.71933.91>.
49. Trost SG, McIver KL, Pate RR. Conducting accelerometer-based activity assessments in field-based research. *Med Sci Sports Exerc.* 2005;37(11 Suppl):S531–43. <https://doi.org/10.1249/01.mss.0000185657.86065.98>.
50. ActiLife 6. <http://www.actigraphcorp.com/product-category/software/>. Accessed 31 Aug 2021.
51. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of accelerometer wear and nonwear time classification algorithm. *Med Sci Sports Exerc.* 2011;43(2):357–64. <https://doi.org/10.1249/MSS.0b013e3181ed61a3>.
52. Schuck P. Assessing reproducibility for interval data in health-related quality of life questionnaires: which coefficient should be used? *Qual Life Res.* 2004;13(3):571–86. <https://doi.org/10.1023/B:QURE.0000021318.92272.2a>.
53. Terwee CB, Mokkink LB, van Poppel MN, Chinapaw MJ, van Mechelen W, de Vet HC. Qualitative attributes and measurement properties of physical activity questionnaires: a checklist. *Sports Med.* 2010;40(7):525–37. <https://doi.org/10.2165/11531370-000000000-00000>.
54. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Int J Nurs Stud.* 1996;33(6):1071–8. <https://doi.org/10.1016/j.ijnurstu.2009.10.001>.
55. Schmidt ME, Steindorf K. Statistical methods for the validation of questionnaires—discrepancy between theory and practice. *Methods Inf Med.* 2006;45(4):409–13. <https://doi.org/10.1055/s-0038-1634096>.
56. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159–74. <https://doi.org/10.2307/2529310>.
57. Svensson E. Ordinal invariant measures for individual and group changes in ordered categorical data. *Stat Med.* 1998;17(24):2923–36.
58. SPSS software. <http://www-01.ibm.com/software/analytics/spss/>. Accessed 31 Aug 2021.
59. SAS software. http://www.sas.com/en_us/home.html. Accessed 31 Aug 2021.
60. Avdic A, Svensson E. Svensson's method 1.1 ed. Interactive software supporting Svensson's method. <http://avdic.se/svenssonsmetod.html> (2010). Accessed 31 Aug 2021.
61. Bolszak S, Casartelli NC, Impellizzeri FM, Maffuletti NA. Validity and reproducibility of the physical activity scale for the elderly (PASE) questionnaire for the measurement of the physical activity level in patients after total knee arthroplasty. *BMC Musculoskelet Disord.* 2014;15:46. <https://doi.org/10.1186/1471-2474-15-46>.
62. Liu RD, Buffart LM, Kersten MJ, Spiering M, Brug J, van Mechelen W, et al. Psychometric properties of two physical activity questionnaires, the AQuAA and the PASE, in cancer patients. *BMC Med Res Methodol.* 2011;11:30. <https://doi.org/10.1186/1471-2288-11-30>.
63. Svege I, Kolle E, Risberg MA. Reliability and validity of the physical activity scale for the elderly (PASE) in patients with hip osteoarthritis. *BMC Musculoskelet Disord.* 2012;13:26. <https://doi.org/10.1186/1471-2474-13-26>.
64. Bowling A. Just one question: if one question works, why ask several? *J Epidemiol Community Health.* 2005;59(5):342. <https://doi.org/10.1136/jech.2004.021204>.
65. Mantri S, Fullard ME, Duda JE, Morley JF. Physical activity in early Parkinson disease. *J Parkinsons Dis.* 2018;8(1):107–11. <https://doi.org/10.3233/JPD-171218>.
66. Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. *Trials.* 2012;13:132. <https://doi.org/10.1186/1745-6215-13-132>.
67. van Poppel MN, Chinapaw MJ, Mokkink LB, van Mechelen W, Terwee CB. Physical activity questionnaires for adults: a systematic review of measurement properties. *Sports Med.* 2010;40(7):565–600. <https://doi.org/10.2165/11531930-000000000-00000>.
68. Forsen L, Loland NW, Vuillemin A, Chinapaw MJ, van Poppel MN, Mokkink LB, et al. Self-administered physical activity questionnaires for the elderly: a systematic review of measurement properties. *Sports Med.* 2010;40(7):601–23. <https://doi.org/10.2165/11531350-000000000-00000>.
69. Baiano C, Barone P, Trojano L, Santangelo G. Prevalence and clinical aspects of mild cognitive impairment in Parkinson's disease: a meta-analysis. *Mov Disord.* 2020;35(1):45–54. <https://doi.org/10.1002/mds.27902>.
70. Food US, Administration D. Guidance for industry. Patient-reported outcomes measures: use in medical product development to support labeling claims. Washington, DC: US Department of Health and Human Services; 2009.
71. Bergman P. The number of repeated observations needed to estimate the habitual physical activity of an individual to a given level of precision. *PLoS One.* 2018;13(2):e0192117. <https://doi.org/10.1371/journal.pone.0192117>.
72. Pedersen ES, Danquah IH, Petersen CB, Tolstrup JS. Intra-individual variability in day-to-day and month-to-month measurements of physical activity and sedentary behaviour at work and in leisure-time among

Danish adults. *BMC Public Health*. 2016;16(1):1222. <https://doi.org/10.1186/s12889-016-3890-3>.

73. Turrisi TB, Bittel KM, West AB, Hojjatinia S, Hojjatinia S, Mama SK, et al. Seasons, weather, and device-measured movement behaviors: a scoping review from 2006 to 2020. *Int J Behav Nutr Phys Act*. 2021;18(1):24. <https://doi.org/10.1186/s12966-021-01091-1>.
74. Helmerhorst HJ, Brage S, Warren J, Besson H, Ekelund U. A systematic review of reliability and objective criterion-related validity of physical

activity questionnaires. *Int J Behav Nutr Phys Act*. 2012;9:103. <https://doi.org/10.1186/1479-5868-9-103>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

